JPRS: 18,406

28 March 1963

OTS: 63-21442

FOREIGN DEVELOPMENTS IN MACHINE TRANSLATION AND INFORMATION PROCESSING

No. 116

U. S. DEPARTMENT OF COMMERCE
OFFICE OF TECHNICAL SERVICES
JOINT PUBLICATIONS RESEARCH SERVICE
Building T-30
Ohio Dr. and Independence Ave., S.W.
Washington 25, D. C.

Price: $1.00

FOREWORD

This publication was prepared under contract for the Joint Publications Research Service as a translation or foreign-language research service to the various federal government departments.

The contents of this material in no way represent the policies, views or attitudes of the U. S. Government or of the parties to any distribution arrangement.

PROCUREMENT OF JPRS REPORTS

All JPRS reports may be ordered from the Office of Technical Services. Reports published prior to 1 February 1963 can be provided, for the most part, only in photocopy (xerox). Those published after 1 February 1963 will be provided in printed form.

Details on special subscription arrangements for JPRS social science reports will be provided upon request.

No cumulative subject index or catalog of all JPRS reports has been compiled.

All JPRS reports are listed in the Monthly Catalog of U. S. Government Publications, available on subscription at $4.50 per year ($6.00 foreign), including an annual index, from the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C.

All JPRS scientific and technical reports are cataloged and subject-indexed in Technical Translations, published semimonthly by the Office of Technical Services, and also available on subscription ($12.00 per year domestic, $16.00 foreign) from the Superintendent of Documents. Semiannual indexes to Technical Translations are available at additional cost.

JPRS: 18,406

FOREIGN DEVELOPMENTS IN MACHINE TRANSLATION AND INFORMATION PROCESSING

No. 116

This serial publication contains selected translations from foreign-language literature on developments in the following fields of language data processing: machine translation studies; questions on structural linguistics, phonological theory, investigation of morphological models, development of syntactic structures and transform analysis; theory of language communication; logical and linguistic problems of constructing information machines and information languages; logical semantics; mathematical and applied linguistics; automatic programming; organization, storage and retrieval of information; automatic recognition; automatic speech input; documentation, etc. This series is published as an aid to U. S. Government research. Report no. 116 deals with

~~Table of Contents~~ ~~Page~~

A LEARNING SYSTEM WHICH UTILIZES STATISTICAL DECISION FUNCTIONS

[Following is a translation of an article by S. F. Kozubovs'kyy, in the Ukrainian-language periodical Avtomatyka (Automation), Vol 7, No 6, Institute of Electrical Engineering, Academy of Sciences, Ukrainian SSR, Kiev, 1962, pages 60-64. This article was submitted to the editors on 28 April 1962.]

The vigorous development of cybernetics in the last decade has been accompanied by intensive use of the most diverse mathematical tools for working out and investigating automatic control systems. The theory of statistical decisions as one of the subdivisions of probability theory has been utilized successfully in the solution of cybernetic problems along with information theory, mathematical logic, and the theory of games.

One of the papers presented at the winter session of the American Institute of Electrical Engineers (23 January to 2 February 1962 in New York) was devoted to the application of the theory of statistical decisions in the analysis and synthesis of cognitive systems which can learn. The paper was read by a worker of the Laboratory of Control Systems and Information, Purdue University (Lafayette), K. S. Fu. It had the title "A Learning System Using

Statistical Decision Functions". [See Note].

([Note] K. S. Fu, "A Learning System Using Statistical Decision Functions," presented at the AIEE Winter General Meeting in New York January 28 - February 2, 1962.)

The paper contained a discussion of questions connected with the analysis and synthesis of a learning system which processes input data in the form of input signals (input measurements) and so-called "learning observations". The system worked as a cognitive device. The problem of recognition was formulated in terms of the theory of statistical decisions. It was shown in the paper that the system can improve its work (learn) by increasing the number of learning observations.

A summary of this paper is given below.

Introduction

The word learning is understood to be some systematic behavior of a system which leads to certain forms of stability [1]. A system which systematically varies its behavior in accordance with a set of input signals can be regarded as a system which learns. In case the system is well designed its behavior improves in accordance with one or more criteria. These criteria are usually linked with decreases in errors or with increases in the stability of the system.

A model of learning from encouragement was studied in different forms. Systems which learn or realize such a model are usually trained by an external "trainer". This trainer selects and applies the encouragement factor on the basis of investigations of the reactions of the system. However, more autonomous learning systems are necessary in which variation of their behavior will be achieved by the system itself, without any external intervention.

In studying adaptive control sysems we start with the idea that they are capable of varying their behavior to correspond with changes in the characteristics of the controlled process which cannot be foreseen or guessed [2]. In this connection an adaptive system is capable of varying its behavior, for example, by changing its imp[illegible] transfer function in such a way as automatically to recognize a signal which appears at random and is subject to noise [3].

When machines for pattern recognition are to recognize new patterns, it is usually desirable to make use of past experience. One of the possible approaches to solving such problems is to design a system which learns from experience to vary its behavior in accordance with changes in the

input signal.

The learning process was illustrated in the paper by means of a system which operated as a cognitive or classifying device. It is clear that a machine which is capable of establishing similarity in input signals solve many problems that must now be solved by human beings. Automatic classification is determination of the class of signals at the input of the device on the basis of a set of measurements of the input signal and a sequence of learning observations. Learning observations are understood to mean measurements of a finite number of possible input signals of each class. After an input signal is recognized, a certain action is performed, for example, a change in the system parameters in order to increase its stability. The losses which the machine incurs as a result of each possible decision are evaluated with the aid of an inseparable weighting function. The purpose of the person who receives the decision is to minimize the probability of losses in the course of recognition.

The analysis and synthesis of a learning system is also regarded as proceeding on the basis of data obtained by measuring the input quantities and the learning observations. The losses which are unavoidable in obtaining a sufficient number of accurate measurements, the abstract

nature of the properties which define the class of the input signal to the system, and the statistical features which are characteristic of measurements make it essential to utilize the so-called statistical methods for analyzing and synthesizing systems.

Although a system is designed for recognition of a certain class of input signals in order to perform a certain action such as varying the behavior of the system after acceptance of the decision, in general, however, the measurements which are to serve as the basis for recognition do not provide enough information for the class of input signal to be recognized successfully at all times. The statistical methods which are applied in such situations constitute the subject of the theory of statistical decision functions [4 - 7].

The Problem of Accepting Statistical Decisions

The problem of accepting a statistical decision is formulated in a fairly general form which includes: 1) the state of a priori knowledge; 2) the losses connected with a possibly incorrect decision; 3) losses due to experimentation. The problem is regarded as connected with a sequence of values of some random variable, let us say, $x_1, x_2, \ldots, x_n$ denoted as $\{x\}$ which can be

considered the result of a series of measurements of the input signal which is to be recognized.

Let Ω be a parameter space. It is subdivided into zones $\omega_1, \omega_2, \ldots, \omega_n$ where ω_i corresponds to the class of the correct signal at the ith input. Here m is the number of possible classes of input signals which must be recognized. Each point in the parameter space corresponds to a set of values of the parameters $\{\Theta\}$. We shall use H_i to denote the hypothesis according to which some $\{\Theta\}$ lies in ω_i. Each set of possible values of $\{x\}$ corresponds to a point in an n-dimensinnal sample space Δ. The Δ space in turn is subdivided into zones $\delta_1, \delta_2, \ldots,$ $\ldots, \delta_m$; here the hypothesis H_i must be accepted only in case $\{x\}$ is contained in δ_1. The problem of accepting a statistical decision must now be defined as the problem of selecting zones $\delta_1, \delta_2, \ldots, \delta_m$ in the sample space for given zones $\omega_1, \omega_2, \ldots, \omega_m$ in the parameter space Ω.

Let $P = (P_1, P_2, \ldots, P_m)$ be an a priori distribution defined on the Ω space. P_i is the a priori probability that the input signal will belong to class ω_i. The decision space which is accepted by the system for recognition is composed of m possible decisions $d_1, d_2, \ldots, d_m$ where d_i is the decision for which

hypothesis H_i is accepted. The weighting function $L(\omega_i, d_j)$ is that loss incurred by the system in case a decision d_j is accepted while the input signal actually belongs to the class ω_i. In case there is a nonsequential process of accepting the decision (in case some correct d is accepted as the decision) the conditional risk in recognizing input signals that belong to class ω_i can be written in the form

$$r(\omega_i, d) = \int_{\Delta} L(\omega_i, d)\, p(x/\omega_i)\, dx, \tag{1}$$

where integration is over the entire Δ space, $p(x, \omega_i)$ is the conditional probability of x on condition that ω_i is the corresponding true class of the measured input signal. For a given P the mean risk is equal to

$$R(P, d) = \sum_{i=1}^{m} P_i r(\omega_i, d) = \tag{2}$$

$$= \int_{\Delta} p(x)\, r_x(P, d)\, dx, \tag{3}$$

where

$$r_x(P, d) = \frac{\sum_{i=1}^{m} L(\omega_i, d) P_i p(x/\omega_i)}{p(x)} \tag{4}$$

is defined as the a priori conditional risk of the correct decision d for the given measurement x.

After that it is necessary to choose the correct decision so as to minimize the mean risk $R(P, d)$ or to minimize the maximum of the conditional risk $r(\omega_j, d)$. The decision rule which minimizes the mean risk is called the Bayes' rule. According to (3), it is sufficient to consider each x separately and to minimize $r_x(P, d)$. In case d' is Bayes' rule, then

$$r_x(P, d') \leqslant r_x(P, d), \tag{5}$$

that is,

$$\sum_{i=1}^{m} L(\omega_i, d')\, P_i p(x/\omega_i) \leqslant \sum_{i=1}^{m} L(\omega_i, d)\, P_i p(x/\omega_i).$$

For a constant function the losses

$$L(\omega_i, d_j) = \begin{cases} 0, & i = j \\ v, & i \neq j \end{cases} \tag{7}$$

and, without losing generality, when $v = 1$, it is easy to show that the risk function 6 is essentially the probability of incorrect recognition. In order to minimize the probability of incorrect recognition it is necessary

to choose d_i on the condition that

$$P_i p(x/\omega_i) > P_j p(x/\omega_j) \qquad \text{for all} \quad i \neq j. \tag{8}$$

In case we define

$$\lambda_{ij} = \frac{p(x/\omega_i)}{p(x/\omega_j)} \tag{9}$$

as the similarity relation, then Bayes' rule for the decision state d_i, that is, H_i is accepted in case that

$$\lambda_{ij} > \frac{P_j}{P_i} \qquad \text{for all} \; i \neq j. \tag{10}$$

The Learning Process in the Recognition Problem

In the preceding section we examined the problem of recognition without learning, with the aid of the theory of statistical decisions. In that section we examined the development of a method for recognizing a class of input signals based on a set of measurements of the input signal and a sequence of learning observations. The process consisted of a set of measurements of the input signal $x = \{x_1, x_2, \ldots, x_n\}$ to be recognized and a set of measurements ℓ_1 of possible input signals of class ω_1

for each $i = 1, 2, \ldots, m$. The last measurements which are called learning measurements are represented in the form $m = \{x_{1,1}, x_{1,2}, \ldots, x_{1,l_1}, \ldots, x_{m,l_m}\}$. The rule for acceptance of decision d is a function which transforms each set of measurements (x, m) into an action which follows recognition of the class of the measured input signal.

Let

$$m = (m_1, m_2, \ldots, m_m), \tag{11}$$

where $m_i = \{x_{i,1}, x_{i,2}, \ldots, x_{i,l_i}\}$ is a set which is composed of l_i training measurements of class ω_i while

$$l = (l_1, l_2, \ldots, l_m)$$

denotes a set with elements $l_1, l_2, \ldots, l_m$. The rule for accepting the decision on the basis of measurements (x, m) is denoted by $d(x, m)$.

For any decision rule d the conditional risk in recognizing input signals which belong to class ω_i is equal to

$$r_l(\omega_i, d) = \int_{\mathcal{M}} \int_{X} L(\omega_i, d)\, p(x, m/\omega_i)\, dx\, dm. \tag{12}$$

where Γ is the space of learning measurements m; $p(x, m/\omega_i)$ is the conditional probability (x, m) in case ω_i is the true class of the measured input signal. In case P is an a priori distribution, the mean risk is equal to

$$R_l(P, d) = \sum_{i=1}^{m} P_i r_i(\omega_i, d) = \tag{14}$$

$$= \int_\Delta \int_\Gamma p(x, m)\, r(x, m)(P, d)\, dx\, dm, \tag{15}$$

where

$$r_{(x, m)}(P, d) = \frac{\sum_{i=1}^{m} L(\omega_i, d) P_i p(x, m/\omega_i)}{p(x, m)} \tag{16}$$

is defined as the conditional risk of the decision rule d for the given measurements (x, m).

If P is fixed but unknown, and if the learning measurements of different classes is statistically unreliable, then

$$P_i p(x, m/\omega_i) = p(x, m, \omega_i) = p(x/\omega_i, m_i) P(\omega_i/m) P(m). \tag{17}$$

Now, equation (14) can be represented in the form

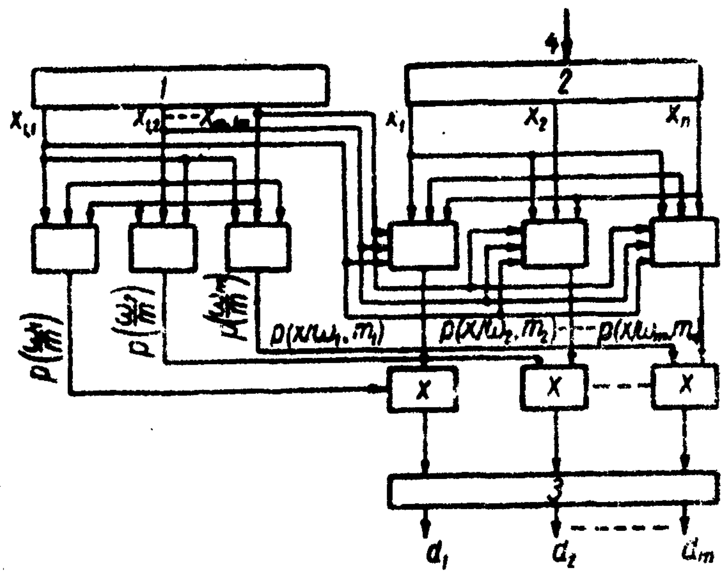
$$R_l(P, d) = \int_\Delta \int_\Gamma p(m) \sum_{i=1}^{m} L(\omega_i, d)\, p(x/\omega_i, m_i)\, P(\omega_i/m)\, dx\, dm. \tag{18}$$

If d' is Bayesian, then we obtain

$$R_l(P, d') \leq R_l(P, d) \tag{19}$$

$$\sum_{i=1}^{m} L(\omega_i, d')\, p(x/\omega_i, m_i)\, P(\omega_i/m) \leq \sum_{i=1}^{m} L(\omega_i, d)\, p(x/\omega_i, m_i)\, P(\omega_i/m) \tag{20}$$

for each set of measurements (x, m) and each decision rule d. The expressions $p(x/\omega_1, m_1)$ and $P(\omega_1/m)$ will be defined on the basis of learning measurements.



A functional diagram of a learning system: 1 - learning measurements m; 2 - measurements x; 3 - determinations of maximum amplitude; 4 - input signal which is recognized.

In order to show that a system can learn or change its characteristics in accordance with learning observations, it is necessary to prove that the Bayesian risk

represented by equation (18) is a nonincreasing function of the number of learning observations. In other words, adding to the learning observations does not increase the risk in the system.

We have arrived at the following theorem: <u>if $l_1^* \geq l_1$, $l_2^* \geq l_2$, ..., $l_m^* \geq l_m$, then for any a priori distribution of</u> P <u>and the Bayesian decision rule</u> d' <u>and</u> d'*

$$R_{l^*}(P, d'^*) < R_l(P, d'), \tag{21}$$

<u>where</u> d'* <u>is the Bayes rule for learning observations</u> l^*.

The theorem can be proved as follows. According to the definition

$$R_{l^*}(P, d'^*) < R_{l^*}(P, d').$$

In accordance with (18)

$$R_{l^*}(P, d') = \int_\Lambda \int_{l^*} p(m^*) \sum_{i=1}^{m} L(\omega_i, d') p(x/\omega_i, m_i^*) P(\omega_i/m^*)\, dx\, dm^*. \tag{23}$$

In view of the fact that d' is not a function of the difference

$$m^* - m = (x_{1l_1+1}, \ldots, x_{1, l_1^*}, \ldots, x_{ml_m+1}, x_{m, l_m^*}), \tag{24}$$

equation (23) can be written in the form

$$R_{l*}(P, d') = \int_{\Delta}\int_{\Gamma}\sum_{l=1}^{m} L(\omega_l, d')\left[\int_{\Gamma^*-\Gamma} p(m^*)\, p(x/\omega_l, m_l^*)\, P(\omega_l/m^*)\, dx_{1, l, l+1 \ldots}\right] \times$$

$$\times\, dx\, dm = \int_{\Delta}\int_{\Gamma} p(m) \sum_{l=1}^{m} L(\omega_l, d')\, p(x/\omega_l, m_l)\, P(\omega_l/m)\, dx\, dm = R_l(P, d').$$

Consequently,

$$R_{l*}(P, d'^*) < R_{l*}(P, d') = R_l(P, d'). \qquad (24)$$

For a constant function the losses

$$L(\omega_i, d_j) = \begin{cases} 0, & i = j \\ 1, & i \neq j. \end{cases}$$

For learning observations Bayes' rule becomes d_i

$$P(\omega_l/m)\, p(x/\omega_l, m_l) > P(\omega_j/m)\, p(x/\omega_j, m_j) \qquad \text{for all } j \neq l \qquad (25)$$

A functional diagram of a model which can learn is shown in the figure.

Conclusions

The design of a learning system and the possibility of the system accepting decisions in accordance with external changes and learning observations were examined. After accepting a decision the system makes use of certain steps directed at improving its operation as it accumulates experience.

The problem of recognition or classification is

formulated in terms of the theory of statistical decisions. Bayes' rule for recognition with learning is determined as a function of the input signal and learning measurements. Learning measurements are utilized in calculating conditional probabilities of all unknown information. In case the values of the unknown parameter which characterizes each class may be obtained from learning observations, the Bayes' risk of a learning system approximates the Bayes' risk of a system with a priori knowledge of its general characteristics.

It was shown that the system can improve its operation or learn as more learning observations are provided. Adding to the learning observations does not increase the risk in the system.

Bibliography

1. R. R. Bush and F. Mosteller, Stochastic Models for Learning, Wiley, N. Y., 1955.

2. J. G. Truxal, "Computers in Automatic Control Systems," Proceedings IRE, January 1961, pages 305-312.

3. C. V. Jakowatz, R. L. Shuey, and G. M. White, Adaptive Waveform Recognition, Report 60-RL-2353E, General Electric Research Laboratory, Schenectady, N. Y., September 1960.

4. R. S. I u, "A Sequential Decision Model for Optimum Recognition," Proceedings of the Second Annual Bionics Symposium, Ithaca, N. Y., August 30 - September 1, 1961.

5. J. Tou and K. S. Fu, "A Digital Control Concept in Nervous System Simulation and Suntheses," Third International Congress on Cybernetics, Namur, Belgium, September 11 - 15, 1961.

6. D. Blackwell and M. Girshik, Teoriya igr i statisticheskikh resheniy (The Theory of Games and Statistical Decisions), translated from the English language, IL [State Publishing House for Foreign Literature], Moscow, 1958.

7. A. Wald, Statistical Decision Functions, Wiley, N. Y., 1950.

8. W. P. Tanner, "Mathematical Models in Sensory Reception," Proceedings of the First Bionics Symposium, pages 263 - 286; WADD Technical Report 60-600, 1960.

9. M. Minsky, "On the Road to Artificial Thought," Tr. In-ta radioinzhenerov (Proceedings of the Institute of Radio Engineers), No 1, January 1961 (Russian translation).

10. O. H. Mowrer, Learning Theory and the Symbolic Processes, Wiley, N. Y., 1960.

5809
CSO: 1801-D

THE ADVANTAGE OF ONE-LAYER COGNITIVE

LEARNING SYSTEMS

[Following is the translation of an article by O. G. Ivakhnenko, Kiev, in the Ukrainian-language periodical Avtomatyka (Automation), Vol 7, No 6, Publishing House of the Academy of Sciences, Ukrainian SSR, Kiev, 1962, pages 10-19. The article was submitted to the editors 12 June 1962.]

Systems Perceiving Letters, Syllables, Words, and Phrases

The previously described system (Alpha) perceives (reads) individual letters, figures, or any other symbols. When letters are projected in chronological sequence on the input of the system it is not necessary to reproduce them exactly as they appear in print, nor is it necessary to enunciate them as they are pronounced individually, because it will produce distortions of the normal speech due to the fact that the correct pronunciation of each letter varies, depending on its position among other letters in accordance with phonetic rules. Thus, it appears that systems perceiving at once entire words or even phrases would be much more effective.

Amalgamation of Certain Outputs of the System by Introduction of an Element of Logical "Or" (Alternate).

When a certain letter is projected on the input of the "Alpha" system, it produces a tension on the output of the corresponding group of associated cells (already informed). This tension is close to its maximum potential: 45, 60, 55, 50, 60, etc. There are certain, comparatively small, aberrations occurring because of the so-called "noises" produced by the differences in the design of a given letter. This happens because the voltage indicator does not respond to these small variations and therefore the entire system tends towards "generalization" or "extrapolation". Thus, the various designs of the same letter acquire the same nomenclature. From this it does not follow, however, that different designs of the same letter could be combined to form one conception. For instance, large print of the letter "E" is entirely different from the small letter "e", and in order to identify each it would be necessary to use diverse groups of associated cells. Should "Alpha" system be adapted to a sound-reproduction device, then both variations

of this letter ("E" and "e") could be combined by using a logical supposition of an alternative "or". In that case the sound reproduction for both would have an identical sound of "e". The same process takes place also in the human mind.

Optimum Division of the Input Information

The size of the system could be determined by the number of the associated cells (relays) required for its functioning. In turn the number of cells depends on the volume of information "read" into the system and also on the methods of its distribution. Thus, larger volume of information predicates a larger size of system. Systems capable of perceiving syllables (reading and sequencing them chronologically) should be larger in size than systems reading only letters. Systems perceiving or reading a complete word would be still larger, and the largest of all would be a system capable of reading complete phrases.

At this point we wish to note that this very fact reveals a potential possibility for eventual design of systems that would be more accomplished than the human brain, - systems that would perceive information by complete sentences or even by entire pages of text, - a feat impossible for the human brain. Now, let us analyse what constitutes an optimum division of information from the point of view of the system's size. In practical application this results in the choice of the number of layers of associating cells.

Syllables are formed with letters, words - with syllables, phrases - with words. Thus, it is possible to design a four-layer system: first layer, of associating cells, will perceive letters, second - syllables, third - words, and fourth - the entire phrase. In designing the system it would be possible to "omit" some of the layers and thus construct a system with three layers, two layers, or even one layer.

Calculations proved that at a given volume of input information, a one-layer system is the most economical, - it requires a minimum of associating cells.

Example

Let us investigate the problem of optimum number of layers in practical application.

Assuming it is required to construct a learning system capable of cognition, as well as reaction to phrases composed of words included in the following sixteen-word vocabulary:

быстро	[fast]	медленно	[slow]
вверх	[up]	назад	[backwards]
вниз	[down]	очень	[very]
влево	[left]	поднять	[lift]
вперед	[forward]	скорость	[speed]
вправо	[right]	снизить	[lowering]
груз	[load]	стоп	[stop]
двигать	[move]	увеличить	[increase]

Let us also assume that a system would be optimum in size if it had a minimum of associating cells. In making our comparison we shall make the following assumptions: a) No more than 12 symbols are necessary to identify an individual letter; b) the longest word in the vocabulary is увеличить [uvelichit'; increase] and it has nine letters and four syllables; c) all words are composed from an alphabet of 24 letters (in Russian); d) vocabulary consists of the sixteen above mentioned words; e) the longest phrase consists of four words, and the number of informative phrases is 20 (only 20 informative phrases are read into the input); f) the number of syllables in the vocabulary is 30.

Let us proceed now with calculations of the number of cells required for different variations of the system in order to determine the optimum number of layers from the point of view of the size of the system. The number of cells in each layer is determined according to the formula:

$$N = nMm$$

where n - is the number of symbols used depending on the system's capacity for division; M - factor of all possible separate images; m - factor of all utilized separate images (there is a number of images that are perceived simultaneously). For instance, for one layer of system (Alpha) reading consecutively 33 letters of the alphabet, we obtain the following: - N = 12 X 33 X 1 = 396 cells.

<u>Systems That do not Store the Perceived Information</u>

For the purpose of brevity we shall limit ourselves to the construction of a system that reads only one word. Figure 1 is a diagram of a two-layer system capable of reading first the letters and then

words. A number of cells in this system equals

1 layer	12 X 24 X 9 = 2592
2 layer	216 X 16 X 1 = 3456
	6048 cells

The number of cells in each layer is determined according to the formula: $N = nMm$ where n - is the number of symbols used depending on the system's capacity for division; M - factor of all possible separate images; m - factor of all utilized separate images (there is a number of images that are perceived simultaneously). For instance, for one layer of system (Alpha) reading consecutively 33 letters of the alphabet, we obtain the following: -$N = 12 \times 33 \times 1 = 396$ cells.

Systems That do not Store the Perceived Information

For the purpose of brevity we shall limit ourselves to the construction of a system that reads only one word. Figure 1 is a diagram of a two-layer system capable of reading first the letters and then words. A number of cells in this system equals

1 layer		12 X 24 X 9 = 2592
2 layer		216 X 16 X 1 = 3456
	Total	6048 cells

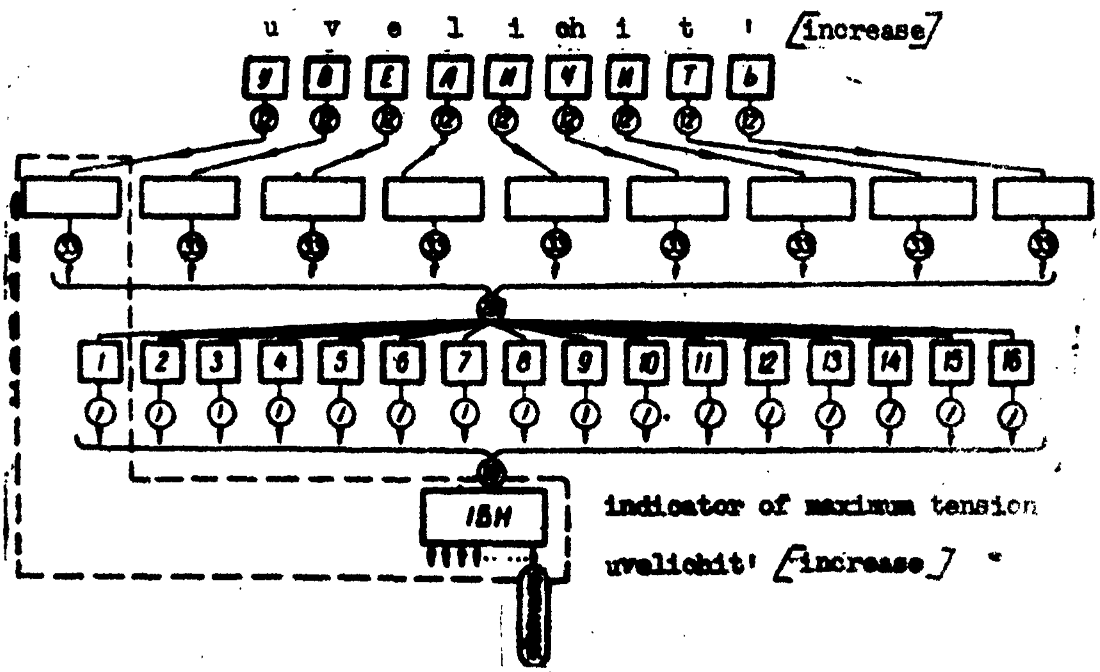


Figure 1. Two-layer system perceiving first letters then words. [The numbers 1-16 designate the words similarly numbered on page 19.]

Figure 2 is another variant of a two-layer system perceiving syllables and then words. The number of cells in this case is:

1 layer	60 X 30 X 4 = 7200	
2 layer	120 X 16 X 1 = 1920	
	Total	9120 cells

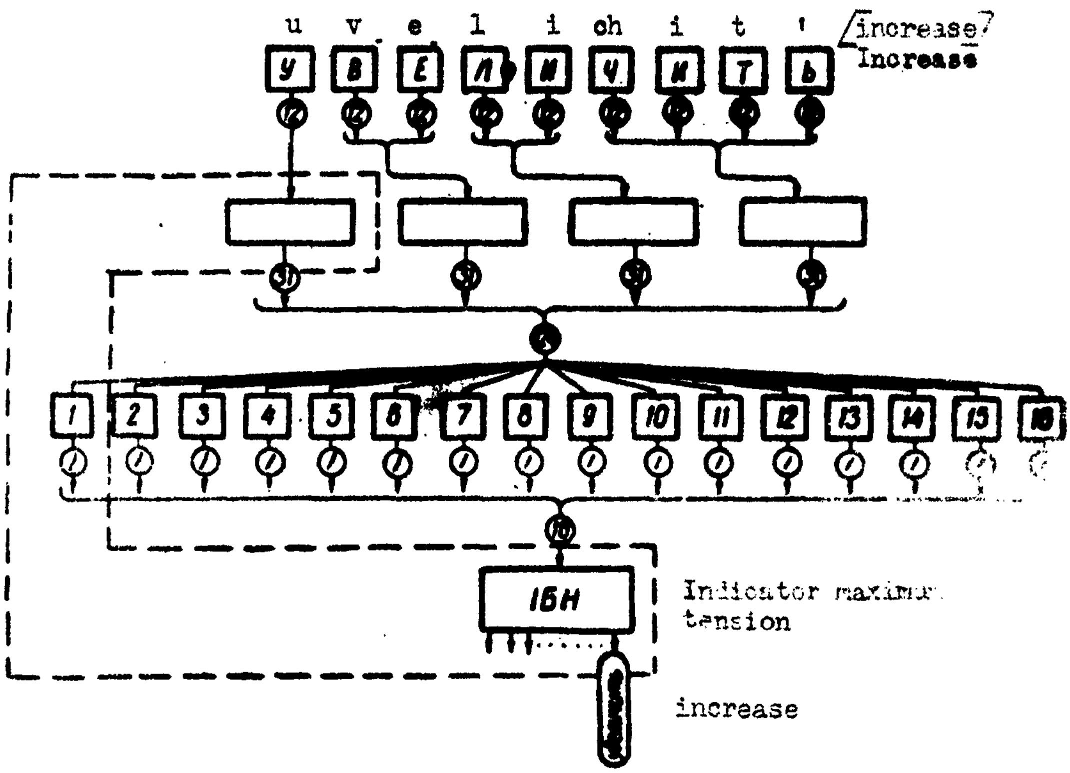


Figure 2. A two-layer system that reads first syllables then words

Designations are the same as in Figure 1.

Finally, Figure 3 is a one-layer system that reads the entire word at once. This system uses

108 X 16 X 1 = 1728 cells

This proves that the one-layer system is in fact an optimum system from the point of view of size (providing the system is not required to store the information).

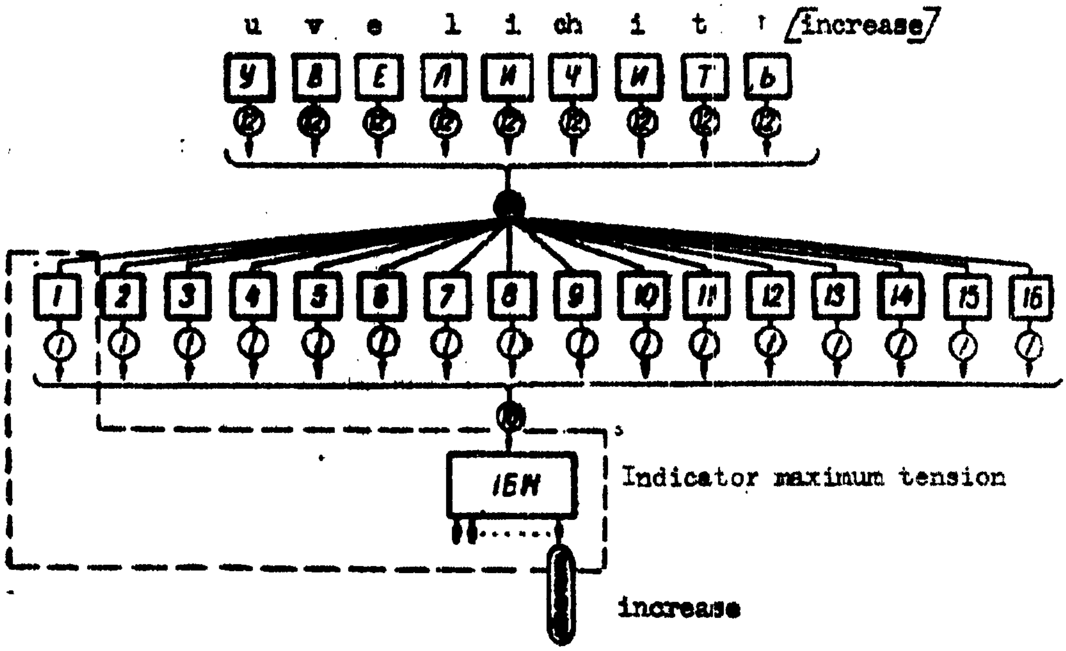


Figure 3. One-layer system which reads the entire word at once.

Designations are the same as in Figure 1.

Let us calculate the number of sells required for a one-layer sys- tem perceiving any of the twenty possible different commands that could be given by using any of the above mentioned sixteen words.

$$N = (108 \times 4) \times 20 \times 1 = 8640 \text{ cells}$$

<u>A System That Could Read Shakespeare</u>

Let us assume that we wish to design a system capable of perceiving all the words in Shakespeare's vocabulary (considering the longest word to be of 10 letters, and the vocabulary to be of 30,000 words). Such case would require

$$120 \times 30{,}000 \times 1 = 3{,}6 \cdot 10^6 \text{ cells}$$

In contrast to a man this system is capable of reading an entire phrase at once which makes it superior to the human brain.

In order to discriminate between all the phrases that could be used in instructing people and adding to it all the possible variations

of them (such as grammatical, for instance) it would be necessary to provide this system with many more cells than are available in the human brain. But this system then, unlike man, would be capable of reading entire phrases at once (and not read by syllables or words as man does). This explains the fact that the science of bionomics, at the present time, is trying to solve the problem of producing inexpensive binary micro elements and develop methods for designing systems with total number of associating cells in the order of 10^{10}.

Systems with Capacity for Storing Perceived Information

What are the methods for reducing the number of required cells? These methods are suggested to us by nature itself. To begin with the reading systems with any number of layers of associating cells could not read the entire information at once; they read progressively, in parts, and use the short-lived ability of memorizing the text as man does. As is known from the general theory of communications, any alleviations in the process of the input of information tend to simplify the entire system.

Introducing special memorizing devices permits a sharp reduction in the number of associating cells. This fact could easily be proven by examples.

In the "Alpha" system the 33 letters in the alphabet could be read either by using 33 groups of associating cells, acting in parallel (as we have demonstrated previously) or by using a single group which is continuously learning and relearning (with the help of the first additional and reversible link) letters that are projected on the input of the system. Information thus obtained is transmitted to the memorizing device. Identification of the letters is carried out by means of their comparative analysis in the output group and also in the memorizing device (i.e. with the help of the cut-off reversible device). For instance, if a sequence of symbols [illegible] already projected on the input of the system, has been assigned number "1", then at any repetition of the same sequence it would be stored again under number "1". In a system that perceives sixteen words (Fig. 1, 2, 3) utilization of memory devices also permits use of only one system "Alpha", capable of reading letters in chronological sequence.

In Fig. 1, 2, 3, dotted lines indicate that part of the system, which remains under maximum exploitation of the memorizing device. Each layer retains only one system.

It could be readily proven, that in case of memorizing information, the one layer system also requires a minimum of associating cells. It is also evident that it will work freer and more often than two or three layer systems (because it will require more operations of reading

and memorizing).

In this respect a compromise solution considering not only the size of the system, but also the speed of its operation, is suggested to us by nature itself. It is impossible for a man to read a complete sentence at once, thus, evidently (considering what has been said before about reading individual letters) an optimum system would be a system, that would read syllables or words with a transitory memorizing of the text. In this fashion we would secure a satisfactory solution for optimum volume of information, to be introduced into the system simultaneously. At the same time we would be able to determine the optimum division of this information into parts, having in mind not only the size, but also the speed of the system. However, leaving out the element of speed, and considering only the size, the one-layer system appears to be the most advantageous, regardless of whether memorizing devices are available or not, providing that the entire input of the system is read as a whole and not in parts.

Systems Employing an Element of Probability

Another method of reducing the size of the system is based on the use of an element of probability. We are apt to guess the meaning of the word by looking only at its beginning, or the end (sometimes the middle). In this fashion we also achieve a reduction in the volume of required information, which at the same time results in the reduction of the number of required associating cells. Considering that nature evidently selected an optimum variant in the creation of the human brain (regarding the economy of required elements) we would assume that the best system is that, which reads the text by syllables (memorizing it), and takes into account the human faculty to make assumptions on the basis of only partial information (subconscious mind).

Development of the optimum sizes of systems is only beginning. There could be no doubt that in the years ahead optimum and invariant systems will be designed (i.e. instantaneous cognitive learning systems) that excell the speed and the capacity of the human brain.

Methods for Calculating "Resolving Capacity" of the Cognitive System

One of the least developed problems is that of the choice of numbers of associating cells in each group - n_1.

The larger n_1 is, the bigger the number of identifying symbols is, and the greater the "resolving capacity" of the system is. At the same time it also lessens the possibility of error in redistribution of images. [See Note] To be sure, besides the number of symbols it is

([Note] To be more exact, with the increase in the number of symbols used the "resolving capacity" either grows or remains stable.)

necessary to calculate also the usefulness and the stability of the symbolism. Let us attempt to derive a mathematical evaluation of the "resolving capacity" in order to utilise it in the selection of the number of associating cells for the group n_1. The number of group m equals the number of individual images (for example, the number of letters in the alphabet). Tension at the output of the group k-oi equals

$$\sum = \sum_{l=1}^{n_1} a_k v_l,$$

where $$1<l<n_1, \; 1<k<m.$$

Indicator of maximum tension (voltage indicator) equalizes these tensions in pairs (after a certain order) and it is able to register the maximum of tension only if it is sufficiently different from the tension appearing at the output. Therefore, we evidently shall be in a position to evaluate "the resolving capacity" by the value of R, which depends on the smallest difference between the two highest tensions in output $\Sigma_1, \Sigma_2, \ldots, \Sigma_m$, when taken in pairs, and from the threshold of the worked out indicator of maximum tension, according to formula $R = \frac{\Delta\Sigma_{min}}{U_{av}}$ where U_{av} is a threshold of the worked out output relays. $\Delta\Sigma_{min}$ is the minimum difference, selected from the two largest among the output tensions. When, for example, R = 3, it means that the system has a necessary reserve of "resolving capacity" with the given number of associating cells n_1 and the after effects v_1 i a_k.

Numerical example. We shall designate "resolving capacity" of the system, which (in a learned condition) has the actual after-effects (with the number of cells in the group n = 12).

$$v_A = a_1 = +1+1-1-1-1-1+1-1-1-1+1-1,$$
$$v_B = a_2 = -1-1+1+1+1-1-1-1+1+1+1+1,$$
$$v_B = a_3 = -1-1+1+1+1+1-1-1+1+1+1+1.$$

Using the above formula, we find that in projecting image A, tensions of the group are equaling

$\Sigma_1 = +12$ unit, $\Sigma_2 = -6$ unit, $\Sigma_3 = -2$ unit, $\Delta\Sigma = +14$ unit.

In projecting image B we obtain accordingly

$\Sigma_1 = -6$ unit, $\Sigma_2 = +12$ unit, $\Sigma_3 = 0$ unit, $\Delta\Sigma = +12$ unit.

In projecting image C -

$\Sigma_1 = -2$ unit, $\Sigma_2 = 0$ unit, $\Sigma_3 = +12$ unit, $\Delta\Sigma = +12$ unit.

The smallest of the differences in tensions, registered by the voltage indicator, is equal to $\Delta\Sigma_{min} = 12$ unit.

Assuming the threshold of the worked out indicator to be $U_{av} = 10_{volt}$ one unit of tension is equal to 5 volts and then the "resolving capacity" will be

$$R = \frac{\Delta\Sigma}{U_{av}} = \frac{12 \cdot 5}{10} = 6.$$

The system has a sixfold reserve. Repeating a number of such calculations with the change of the associating cells n_1 it is easy to see that the "resolving capacity" of the system decreases when n_1 becomes smaller. The smaller n_1, the smaller the "resolving capacity" (Fig. 4) When $R < 1$, the system ceases to discern the images (Voltage indicator does not work).

"The resolving capacity" R

Projected		$n_1=12$	$n_1=11$	$n_1=10$	$n_1=9$	$n_1=8$	$n_1=7$	$n_1=6$	$n_1=5$	$n_1=4$	$n_1=3$	$n_1=2$	$n_1=1$
A	Σ_1	+12	+11	+10	+9	+8	+7	+6	+5	+4	+3	+2	+1
	Σ_2	−6	−5	−6	−5	−4	−5	−4	−5	−4	−3	−2	−1
	Σ_3	−2	−1	0	+1	0	+1	+2	+1	0	−1	0	+1
	$\Delta\Sigma$	+14	+12	+10	+8	+8	+6	+4	+4	+4	+4	+2	0
B	Σ_1	−6	−5	−6	−5	−4	−5	−4	−5	−4	−3	−2	−1
	Σ_2	+12	+11	+10	+9	+8	+7	+6	+5	+4	+3	+2	+1
	Σ_3	0	−1	0	−1	0	+1	0	−1	0	+1	0	−1
	$\Delta\Sigma$	+12	+14	+10	+8	+8	+6	+4	+4	+4	+2	+2	0
C	Σ_1	−2	−1	0	+1	0	+1	+2	+1	0	−1	0	+1
	Σ_2	0	+1	0	−1	0	+1	0	−1	0	+1	0	−1
	Σ_3	+12	+11	+10	+9	+8	+7	+6	+5	+4	+3	+2	+1
	$\Delta\Sigma$	+12	+12	+10	+8	+8	+6	+4	+4	+4	+2	+2	0
$\Delta\Sigma_{min}$		+12	+12	+10	+8	+8	+6	+4	+4	+4	+2	+2	0
R		+6	+6	+5	+4	+4	+3	+2	+2	+2	+1	+1	0

These calculations could be used for rational selection of the number of cells in a group in order to secure the given reserve of "resolving capacity" - R for the system. In passing, let us note that another additional link could be achieved in such a way that with a constant number of cells in the group n_1 a selection could be made from the factor of n_2 symbols that would provide greatest "resolving capacity". The prompting in this case should come from the counting device that would determine R in accordance with the above rules.

Figure 4. Approximation of relationship between the "resolving capacity" of the cognitive system and the number of associating cells in a group.

The Principles of Designing Systems for Control of Complex Processes

As is known (2) the present system of control in complex processes could be divided into the following parts:

1) Means and methods of perceiving situation (identification)

2) Logic-taking decision

3) Means and methods of influencing the object (activization)

4) Object of control

These four parts represent a system of control, which could be developed by contemporary theory of control, including the theory of invariants. Dispersion analysis tells us that the resolving the capacity of the system provides the most effective influence on the object. In more simple cases any stabilisation system could be designed completely on the basis of the above given schemes. Control and regulation of the size helps in solving the problem of identification. The logic of it is that the ability to compare the regulated product with its purpose

helps in discovering the cause of deviation. Servomotor and booster help to solve the problem of activisation.

All these factors together with the object of regulation implement a well known system of stabilisation of the regulated quantities. The problem of identification at the new contemporary level is solved by a battery of devices. The above described resolving systems also could be used for solving that problem. In fact, it is not difficult to see that the problem of discerning letters is not essentially different from the problem of discerning other situations that arise, for example, in the industrial production.

Let us give a very simple example. A tanker comes into port for delivery of its cargo of oil. The ship's presence at a given berth in the warf is given a symbol + 1 unit, absence - 1 unit, ship filled to a certain level + 1 unit, below that level - 1 unit, etc. The number of the output reactions would correspond to the number of the possible situations; naturally, some of them could be combined through the use of the logical alternative "OR", and would still produce the desired reaction of the system. We wish to point out again that control systems regarding complex processes, whether in their entirety or in parts, for the purpose of managing an operation, could be carried out as an open system, as a closed system, or as a combination of both.

An Open System of Management consists of an identifying part, a logical decision part, a control part and an object of management not connected with the identifying part. A system of that sort could be only a cognitive learning system, because the logical decision in all its particulars must be given by a man. For each output of the identifying part of the system, a certain reaction should be introduced into the system by a man (in order to influence the object). Therefore, regardless of the fact that the identifying part of the system could have a self-learning capacity, the entire system still could be an open one and lack that magic quality.

Closed System of Management differs from the previous by the fact that its identifying part is closely tied to the object of management. The existence of a closed circuit and a return connection does not per se indicated self-sufficient and self-organised system. Self-organisation (or self-learning) in a closed system takes place only when:

1) it selects and puts in operation the algorisms that couple the identifying part to the possible reactions of the system, and man's part consists only in indicating the purpose of the action;

2) it carries out this selection and improves upon it;

3) under certain conditions it also finds the rational purpose of its action.

Let us analyse these problems of self-organisation one by one:

1. Systems with a Constant Algorism of Optimum Reaction. In order to clarify the action of such systems let us turn again to our oversimplified example: there are two possible berths for the oil-carrying tanker. The controlling system carries out the unloading of the ship. The following situations are possible: a) both berths are free, b) the first berth is occupied, c) the second berth is occupied, d) both berths are occupied. The simplest identifying system should have four outputs corresponding to the number of the situations. Considering the loading and unloading costs, pumping expenses, etc., the number of the possible reactions could be, evidently, much greater than the number of the outputs of the identifying system, because the system could take into account previous history, and react to the combination of outputs. However, the possible maximum number of reactions cannot exceed the total number of possible outputs, and they provide the ultimate figure. Thus all the variants of loading and unloading could be calculated on the counting machines or on the models establishing sequences of reactions (algorisms of action), which establishes the extremes of any index (for instance, minimum of unloading costs) or shows the optimum relationship between a number of indices, (for example, optimum relationship between the speed and the cost of unloading).

The purpose of the action in such a system, i.e. demonstration of extreme or the optimum relationship, is indicated by man and the role of the system consists in calculating and bringing about the optimum method of reaching this purpose by means of computing and comparing all the possible variants.

2. Systems with Self-Improvements in Algorisms for Finding Optimum Reactions. Systems with constant computation of all possible variations on the counting machines are accomplishing only the first problem of self-organisation. Self-organisation of the system could be improved, if we could add the capacity for improvement to it on the basis of its past performance.

It should be remembered that besides a computation and comparison of all the possible variants, we could also introduce statistical research to achieve the first satisfactory result. In such a case computation and comparison concerns only those variants which are closest to the very extreme.

Self-learning systems with statistical research compute the results of the previous performances and test those decisions that appear to be the most promising. They reject unsatisfactory ones and do not seek answers in areas proven to be inappropriate. Self-learning in a system helps it make proper decisions faster and better. Self-learning systems seeking the optimum or the extreme are characterized by automatically improving the algorisms of search for their objectives. This

is achieved by adding another system, of a different order to the existing one, and treating the latter as an object of the second (hierarchy of the extreme systems). Self-learning results in a speedier and better work of the system that commands "the knowledge of experience". The words "faster and better" indicate that with reference to both we are confronted either by the manifestation of statistical approach or by the supplemental return connection which is always present when we deal with the avalanche type process of growth of some physical quantities. In this case action of the supplemental return connection leads to a continuous increase in speed of the system's action and its accuracy.

Thus the second of the three above mentioned problems of self-organization is solved with the help of statistical research or supplemental return connections.

3. Systems with Research Regarding the Purpose of Self-Organization. It is possible to make the system more perfect by adding to it the capacity for searching and finding the goal of self-organization. Let us pose a question here. Is it possible, in principle, to remove man from his ultimate role, the role of establishing the systems' goals or changing them? In principle this, evidently, is quite possible. But then again you have to consider the process of evolutionary struggle, death and survival. Those systems that survived either by chance or f. any other reasons, succeeded in doing so because they have found more appropriate aims and have transmitted them down the line. This is evidently the process that took place in nature, as a result of which we have the existing species of animal and plant life. Both, as is well known, possess well designed management systems.

It is evident that in designing management systems for industrial purposes, knowledge of systems' aims does not create a problem because the aims of management are very clear and could be determined by men without any difficulty. The rest of self-organization could be carried out by the system itself.

Bibliography

1. O. H. Ivakhnenko. "Self-learning systems with supplemental return connections," Avtomatyka (Automation) Vol 3, 1962.

2. E. Mishkin, L. Braun (editors), Adaptive Control Systems, 1961.

6481
CSO: 1801-D